THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 288, NO. 19, pp. 13563–13574, May 10, 2013

Specificity of Processing α-Glucosidase I Is Guided by the Substrate Conformation

CRYSTALLOGRAPHIC AND IN SILICO STUDIES[*] [S]

Received for publication, February 13, 2013, and in revised form, March 22, 2013 Published, JBC Papers in Press, March 27, 2013, DOI 10.1074/jbc.M113.460436

Megan K. Barker[‡1] **and David R. Rose**[‡§2]

From the [‡]*Department of Medical Biophysics, University of Toronto, Ontario Cancer Institute, Princess Margaret Hospital, Toronto, Ontario M5G 2M9 and the* [§]*Department of Biology, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*

Background: The enzyme "GluI" is key to the synthesis of critical glycoproteins in the cell.
Results: We have determined the structure of GluI, and modeled binding with its unique sugar substrate.
Conclusion: The specificity of this interaction derives from a unique conformation of the substrate.
Significance: Understanding the mechanism of the enzyme is of basic importance and relevant to potential development of antiviral inhibitors.

Processing α-glucosidase I (GluI) is a key member of the eukaryotic *N*-glycosylation processing pathway, selectively catalyzing the first glycoprotein trimming step in the endoplasmic reticulum. Inhibition of GluI activity impacts the infectivity of enveloped viruses; however, despite interest in this protein from a structural, enzymatic, and therapeutic standpoint, little is known about its structure and enzymatic mechanism in catalysis of the unique glycan substrate $Glc_3Man_9GlcNAc_2$. The first structural model of eukaryotic GluI is here presented at 2-Å resolution. Two catalytic residues are proposed, mutations of which result in catalytically inactive, properly folded protein. Using Autodocking methods with the known substrate and inhibitors as ligands, including a novel inhibitor characterized in this work, the active site of GluI was mapped. From these results, a model of substrate binding has been formulated, which is most likely conserved in mammalian GluI.

N-Glycosylation, the addition of a glycan to an asparagine residue, is the most common post-translational modification in eukaryotes, with over half of all eukaryotic proteins estimated to be glycosylated (1). The presence and identity of an *N*-glycan on a protein affect stability, folding, and intermolecular interactions. More broadly, *N*-glycans play critical roles in reaction kinetics modulation, intracellular protein trafficking, and cell-cell adhesion and communication. Several enveloped viruses require *N*-glycosylation of their coat proteins for successful infectivity (2); as such, the *N*-glycosylation pathway is of key antiviral therapeutic interest. Thus, structural and mechanistic investigations into enzymes that mediate *N*-glycosylation are of fundamental importance to studies of human health and physiology.

Assembly and processing of the protein-glycan conjugate takes place in the endoplasmic reticulum and Golgi in an intricate, branching, multistep system (3). At the initial stage of this pathway, the enzyme "processing α-glucosidase I" (GluI)[3] catalyzes the selective removal of the terminal glucose from the newly linked glycoprotein, in a co-translational process. GluI holds a key regulatory position in the *N*-glycosylation pathway by maintaining forward momentum of the glycan transfer reaction and by working in conjunction with the folding quality control system (4–6). Loss or inhibition of GluI prohibits further glycoprotein processing in the endoplasmic reticulum, and also has an impact on the resident lipid-linked and free oligosaccharide species' populations (7–9). Of therapeutic relevance, inhibition of GluI activity results in reduced assembly and infectivity of several enveloped viruses including hepatitis B and C, influenza, HIV, and others (10–14). However, the known inhibitors are not specific, resulting in undesirable side effects. A specific inhibitor for GluI could impact viral infectivity, whereas avoiding off-target interactions. The knowledge of the structure and catalytic mechanism of GluI would greatly aid in design or discovery of such an inhibitor.

GluI is a single-pass type II transmembrane protein of ~80–110 kDa, with the bulk of the protein including the catalytic region found in the endoplasmic reticulum lumen (15–18). The biological substrate for GluI is $Glc_3Man_9GlcNAc_2$, whether dolichol-linked, protein-linked, or as a free oligosaccharide; GluI cleaves the terminal glucose-α(1→2)glucose glycoside linkage, releasing glucose. In all eukaryotic homologs tested, GluI is specific for this linkage, and the minimum cleavable substrate is glucotriose with α(1→2) and α(1→3) linkages as found in the native substrate (15, 19–22). The glucose-α(1→2)glucose disaccharide, kojibiose, inhibits GluI activity weakly (15, 23). Interestingly, the only documented biological occurrence of this glucotriose is found in the eukaryotic *N*-glycosylation pathway; thus, the relationship between this enzyme and this substrate is unique in biology.

* This work was supported by Canadian Institutes for Health Research Grant MOP111237, Heart and Stroke Foundation Grant NA-6305, and the Natural Science and Engineering Research Council of Canada (NSERC).
[S] This article contains supplemental Figs. S1–S6 and Tables S1–S3.
The atomic coordinates and structure factors (code 4J5T) have been deposited in the Protein Data Bank (http://wwpdb.org/).
[1] To whom correspondence may be addressed: 610 University Ave., Rm. 7-411, Toronto, Ontario M5G 2M9, Canada. E-mail: megan.barker@utoronto.ca.
[2] To whom correspondence may be addressed. E-mail: david.rose@uwaterloo.ca.

[3] The abbreviations used are: GluI, processing α-glucosidase I; GH, glycoside hydrolase; r.m.s., root mean square; DNJM, deoxynojirimycin; PDB, Protein Data Bank.

As a glycoside hydrolase (GH), GluI is a member of the CAZy database GH family 63 and clan GH-G whose members operate via an inverting mechanism; the catalytic acid and base are not definitively known (24, 25). A substrate binding motif in the rat and mammalian homologs has been proposed (26), but no eukaryotic structures have been determined. Two structures have been solved of prokaryotic GH63: the *Escherichia coli* homolog YgjK (27) (PDB code 3D3I), and the *T. thermophilus* homolog TTHA0978 (PDB ID 2Z07, RIKEN structural genomics). Both structures contain an $(\alpha/\alpha)_6$ toroid fold, whereas YgjK possesses an additional N-terminal super-β-sandwich domain. Neither of these structures is sufficiently similar to mammalian GluI to act as a realistic model at the atomic level.

Much of what we know about the characteristics of GluI has been learned from studying the *Saccharomyces cerevisiae* enzyme, Cwh41p, which we have stably purified from *Pichia pastoris* overexpression as a transmembrane-deletion construct, Cwht1p (28). Cwh41p and human GluI share 24% overall identity and from 34 to 59% identity in the catalytically active C-terminal domain (17), and so similar structures are expected. Yeast and human GluI share similar substrate specificity, pH optimum, and inhibitor sensitivity (19, 29). In both enzymes, arginine, tryptophan, or cysteine modification results in an inactive enzyme (26, 30, 31). Thus, the yeast enzyme serves as a good experimental model to learn more about the structure, substrate specificity, and enzymatic mechanism of human GluI.

In this work, we have determined the structure of Cwht1p to 2 Å. Based on structural similarity, the active site residues are proposed to be a glutamate (Glu^{771}) and an aspartate (Asp^{568}) in the center of the $(\alpha/\alpha)_6$ barrel that forms the catalytic C-terminal region. The crystal packing prohibits experimental active-site investigations due to occlusion of the active site by a His_6 purification tag from a crystal contact. Therefore, the active site was investigated by *in silico* methods using small ligands. These analyses indicate a basis for the substrate specificity of GluI and features important for inhibitor development.

EXPERIMENTAL PROCEDURES

Data Collection, Structure Phasing, and Refinement—The growth of diffracting Cwht1p crystals was described in our previous work (28). A heavy-atom approach to phasing was taken with a panel of heavy atoms first screened using a gel-shift assay to prioritize the compounds (32). The successful phasing signal was obtained from ethyl mercuric phosphate, which was soaked into the crystal at 1 mM for 16 h prior to freezing.

Cwht1p crystals were looped into 1:4 paratone:mineral oil (Hampton) as cryoprotectant and flash-frozen in a nitrogen stream at 100 K. Diffraction data for the native crystals were collected on an ADSC Quantum-4 CCD detector at the Cornell High Energy Synchrotron Source (CHESS), F-1 beamline. Data from the mercury-derivative crystals were collected on the ADSC Quantum-315 CCD detector at the Advanced Photon Source (APS), BioCARS, 14-BM-C beamline, at the Argonne National Laboratories.

All diffraction data were processed using HKL-2000 (33). Single anomalous diffraction phasing was performed with AutoSol within the PHENIX (Python-based Hierarchical Environment for Integrated Xtallography) program package (34). GluI crystallized in space group $P2_12_12_1$, with one molecule in the asymmetric unit. The Matthews coefficient is 2.6 $Å^3$/Da, corresponding to 49.5% solvent. Single anomalous diffraction phasing with the mercury-derivatized dataset gave eight heavy atom sites with a phasing figure of merit 0.493 and Bayes correlation coefficient of 53.3. Following phasing and density modification, an initial model was auto-built using the automated program Autobuild in PHENIX (34), containing 588 residues, with R/R_{free} of 0.341/0.366. This initial model was manually built in Coot and refined with Refmac in CCP4 against the native data set, giving a final model containing 788 (of 813) residues, with an R/R_{free} of 0.204/0.178 (35–37). The Molprobity web server was used to evaluate the structure quality during and after refinement (38). As calculated by MolProbity, there were no Ramachandran outliers, no Cβ deviations, and an acceptable value (<1%) of poor rotamers (0.57%) in this structure. The coordinates of the final refined model have been deposited in the Protein Data Bank (39) with code 4J5T.

Cwht1p Structure Analysis, Comparison, and Prediction Methods—A variety of software packages and web servers were used to evaluate and analyze the Cwht1p structure. Crystal packing interfaces were evaluated using the PDBePISA (Protein Interfaces, Surfaces, and Assemblies) web server (40). Cwht1p was queried using the DALI Lite version 3 server to determine similar structures in the Protein Data Bank for comparative analysis (41). When querying based on domains, the Cwht1p model was split into the N-domain (residues including linker region; residues 1–386) and the C-domain (excluding the non-native C-terminal tag; residues 287–800). Pairwise structural comparisons and r.m.s. deviation calculations were performed using the Dali Pairwise server. Structure-based sequence alignments were performed using the PROMALS3D server (42). Electrostatic surface calculations were performed with CHARMM using the PBEQ server (43). Structural figures were generated using the PyMol program. Two-dimensional interaction diagrams were adapted from the PoseView server output (44).

Construction, Expression, Activity, and Crystallization of Cwht1p Mutants—Catalytic residues were proposed based upon structural similarity with other glycoside hydrolases. These residues (Asp^{568} and Glu^{771}) were mutated using site-directed mutagenesis of Cwht1-pPICZαAΔXhoI (plasmid described previously (28). The primers used for the single (DN, DA, EQ, EA) mutations are as follows (5′ to 3′ forward and reverse primers for each): DN, GCCAAGCGGTATGAACGACTATCCTAGAGCACAAC, GTTGTGCTCTAGGATAGTCGTTCATACCGCTTGGC; DA, GCCAAGCGGTATGGCAGACTATCCTAGAGCACAAC, GTTGTGCTCTAGGATAGTCTGCCATACCGCTTGGC; EA, GGAAGAACAAGGTTATTGTTATGCCAATTACAGTCCGATAGATGG, CCATCTATCGGACTGTAATTGGCATAACAATAACCTTGTTCTTCC; and EQ, GGGAAGAACAAGGTTATTGTTATCAAAATTACAGTCCGATAGATGGTC, GACCATCTATCGGACTGTAATTTTGATAACAATAACCTTGTTCTTCCC.

Double mutants (DAEA, DNEQ) were constructed sequentially, using the single mutants; all mutations were confirmed by sequencing. The plasmid PCR products using these primers were digested by DpnI to remove non-mutated clones, and

transformed into *P. pastoris* X33 using the PEG transformation method as described for native Cwht1p (28). The six single and double mutant proteins were purified and their activity tested using the same protocol as the native Cwht1p (28).

Note that the residue numbering in this work is consistent with our previous Cwht1p publication (28), and begins after the 33 additional residues present in the full-length Cwh41p. Thus, residue 1 is methionine, and the Cwht1p construct expressed here from pPICZαAΔXhoI contains 4 residues prior from the expression vector (Glu −3 to Phe 0). Both the N and C termini are non-native in this protein, being a N-terminal transmembrane-deletion construct with a C-terminal His_6 tag.

For low-resolution structural characterization of the mutants, circular dichroism (CD) was performed. The CD spectra of native and mutant proteins were measured on an Aviv model 62 DSA spectrometer. Samples were prepared with 3.5 μM protein in 20 mM sodium phosphate, pH 6.8, 100 mM NaCl. Spectra were collected from 320 to 205 nm with a 1.0-nm bandwidth and 1-mm path length.

Inhibitor Screening—Michaelis-Menten kinetic parameters have been determined for Cwht1p with synthetic trisaccharide and tetrasaccharide substrates, subregions of the $Glc_3Man_9GlcNAc_2$ biological substrate (28, 45). The same assay was used here to screen a panel of compounds for their inhibition of Cwht1p. Briefly, Cwht1p is incubated with the tetrasaccharide substrate for 10 min, and the reaction subsequently quenched. Product concentration is then determined using the colorimetric Glucose Oxidase assay (Sigma). Dimethyl sulfoxide-dissolved compounds (supplemental Fig. S1) were screened for inhibition by preincubating the compound with the enzyme for 10 min prior to substrate addition. Enzymatic activity was calculated as a relative to the control. The compounds screened for inhibition were initially assayed at 1 mM final concentration; compounds showing substantial inhibition were subsequently tested at a range of concentrations.

Intrinsic Fluorescence of Cwht1p with Glucose—Glucose-Cwht1p binding was investigated using tryptophan fluorescence experiments. Fluorescence spectra were measured with a Shimadzu Scientific Instruments RF-5301PC Spectrofluorophotometer. Samples were excited at 295 nm and the fluorescence emission observed from 290 to 420 nm. The slit width was set to provide a band pass of 10 nm for excitation and 3 nm for emission. The cuvette was thermostatted at 20 °C. Samples were prepared with 15 μM Cwht1p, 20 mM sodium phosphate, pH 6.8, 100 mM NaCl, and a range of glucose concentrations (35 mM to 3.5 M).

Docking Sugars and Inhibitors into the Active Site—Ligands were docked *in silico* into the proposed active site of Cwht1p. The ligands (glucose, miglitol, deoxynojirimycin, kojibiose, and glucotriose) used in these docking runs include sugars that form part of the biological substrate, as well as known inhibitors. All docking runs were performed using AutoDock Vina (46). Energy-minimized oligosaccharide PDB files were generated using the GLYCAM Carbohydrate Builder. Cwht1p and ligand pdbqt files were prepared for docking input in the AutoDock Tools GUI, which calculates surface grids with partial atomic charges. All non-ring bonds in the ligands were set as rotatable. AutoDock Vina was used to run the docking in a 36 × 36 × 36 $Å^3$ box around the active site. The exhaustiveness was set to 150, and the top output poses were ranked by their calculated binding affinities.

RESULTS

Cwht1p Structure—The structure of Cwht1p was solved to 2.04 Å (structure and topology in Fig. 1 and supplemental Fig. S2; crystallographic statistics in supplemental Table S1 and "Experimental Procedures," PDB code 4J5T). Cwht1p is a globular protein, with dimensions roughly 95 × 45 × 55 Å, consisting of two domains. The N-domain (residues −2 to 278) consists of an N-terminal α-helix (NH1), a 13-strand super-β-sandwich (NS1-NS13), and two α-helices (NH1,2) between NS11 and NS12. The N- and C-domains are connected by 42 residues, which includes one linker α-helix, LH1. The C-domain (residues 320 to 808) consists of 12 helices (CH1-CH12) in an $(\alpha/\alpha)_6$ toroid bundle, with an extra structural unit, the C′-region, containing two α-helices (C′H1,2) and eight β strands (C′S1–8). Density was missing for one amino acid at the N terminus, and for two loops (residues 226–231 between NH_2 and NH3, and residues 474–492 between CH4 and CH6). Two *N*-glycans, consisting of two GlcNAc residues each, were visible in the electron density, linked to asparagines 9 and 89 in the N-domain of the structure. One disulfide bond is present between cysteines 636 and 652. The final refined model contains 284 water molecules.

Cwht1p expressed here in *P. pastoris* is an *N*-glycoprotein, as previously determined through PNGase treatment and a band shift on SDS-PAGE (28). Based on the 2–4-kDa band shift upon deglycosylation, and assuming identical composition of both glycans (N9 and N89 in the structure), roughly 3–8 more residues would be likely present on each glycan, but are disordered in the crystal. Neither glycan site is located near a crystal packing interface, and both glycans point away from the protein, with no major protein-sugar contacts seen in the modeled residues. There was no unmodelled density at nearby symmetry-related proteins in the crystal to suggest that glycans were participating in the crystal packing at the distal end.

Crystal Packing and Interfaces—There are two major crystal packing interfaces (Fig. 2*A*), with a calculated free energy of solvation of −14.5 and −5.1 kcal/mol, for the large (1456 $Å^2$) and small (747 $Å^2$) interfaces, respectively, with negative values indicating thermodynamic favorability. Contained within the large interface is the His_6 tag of a symmetry-related mate in the crystal (Fig. 2, *B* and *C*), which itself contributes 491 $Å^2$ of the buried surface area, and −5.4 kcal/mol of the solvation free energy of that interface.

Similar Structures—A DALI search of the PDB gave several hits with high *Z*-scores and good structural similarities to Cwht1p (supplemental Table S2). Within the list of hits, the shorter-length proteins aligned only to the C-domain. The longer proteins possessed regions similar to both the N- and C-domains of Cwht1p, with variability in the interdomain orientation. Notably, the top two DALI hits were YgjK and TTHA00978, the two bacterial GH Family 63 structures. A structure-based multiple sequence alignment of Cwht1p, YgjK, and TTHA00978 was performed in conjunction with human

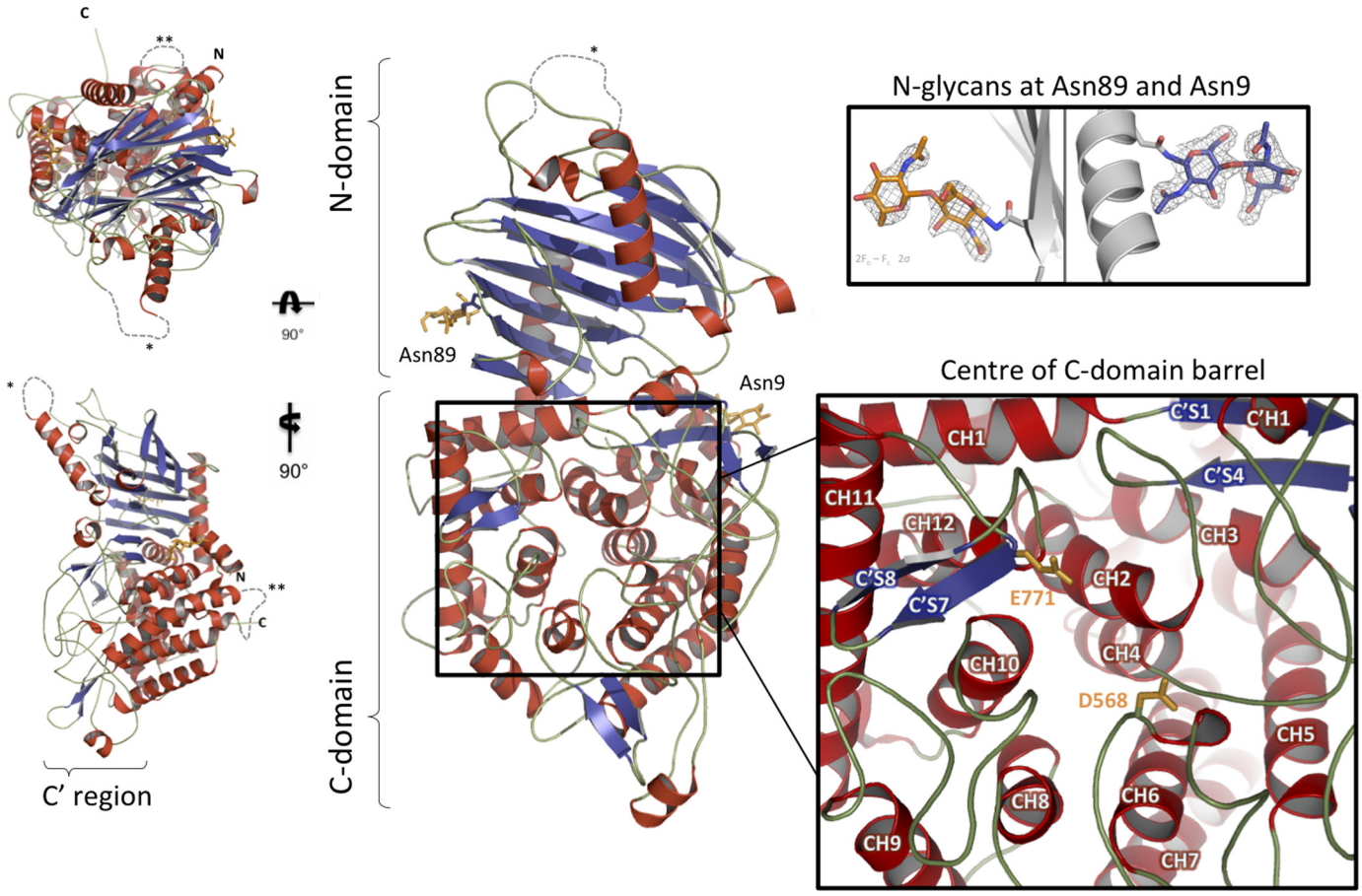

FIGURE 1. **Structure of Cwht1p.** α-Helices are shown in *red*; β-sheets in *purple*; random coil in *green*; glycans shown in *orange*. In the *center view*, the N and C termini are at the *back* of the structure, pointing into the page. Rotated views at *left* are not to scale with the center view. The two loops that could not be modeled due to lack of density are indicated with a *single asterisk* (residues 226–231) and *double asterisks* (residues 474–492). *Top right* shows a stick model of $GlcNAc_2$ linked to each of asparagine residues 89 and 9, modeled into electron density map. At *bottom right*, a larger view of the C-domain center, with secondary structure elements labeled as discussed under "Results" (C or C′ for C-domain or C′-region, respectively; *H* or *S* for the helix or strand secondary structure, respectively, numbered within those regions); side chains of residues Asp^{568} and Glu^{771} are shown in *orange sticks*. (Note: Fig. 4 is in the same reference orientation.)

and rat primary sequences, with the C-terminal alignment presented in supplemental Fig. S3.

GH Family 63 Structures and Alignment—The two domains in Cwht1p are found in numerous other solved protein structures (supplemental Table S2). Cwht1p shares the highest similar overall architecture with the two-domain protein YgjK (*E. coli* GH 63 member); in an overall superposition, with C-domains of YgjK and Cwht1p matched, the N-domains are rotated ~15° relative to each other along the major axis of the protein (supplemental Fig. S4*A*). When comparing individual domains, the N- and C-domains share backbone r.m.s. deviations of 2.6 and 2.5 Å, respectively. TTHA0978 (the *Thermus thermophilus* GH 63 member) has a single domain, structurally similar to the C-domain of Cwht1p (backbone r.m.s. deviations 2.7 Å). No other GH 63 structures have been solved to date. Trehalase (Tre37), the third hit, is a known glycoside hydrolase in family 37, and shares the GH-G fold architecture with GH 63 proteins. Subsequent neighbors are glycoside hydrolase members of the GH-L clan, containing similar $(\alpha/\alpha)_6$ barrel and super-β-sandwich domains. Both TTHA00978 and Tre37 contain only the $(\alpha/\alpha)_6$ toroid domain, whereas the remainder of the hits listed additionally contain the super-β-sandwich domain. Alignment at the known active site between the top four similar structures and the Cwht1p structure is indicated in Table 1 and supplemental Fig. S4*C*.

Mutant Protein Purification and Characterization—Based upon proposed active site residues suggested by the structure, six point mutants of Cwht1p were expressed in *P. pastoris* and purified using nickel-nitrilotriacetic acid IMAC and gel filtration chromatography (supplemental Fig. S5, *A–C*). All mutants show no catalytic activity against the tetrasaccharide substrate (supplemental Fig. S5*D*). Circular dichroism data of the mutants and native protein (supplemental Fig. S5*E*) indicate no gross structural changes between the mutant and native proteins.

Glucose and Inhibitor Binding Experiments—Prior to *in silico* docking experiments, we examined Cwht1p binding with glucose (the reaction product) and candidate inhibitors. Tryptophan fluorescence of Cwht1p in the absence or presence of glucose was performed, with the spectra presented in Fig. 3*A*. No shift in the peak wavelength was seen upon the addition of glucose; λ_{max} was observed at 330 nm in all spectra. Increasing the glucose concentration results in increased intensity of tryptophan fluorescence, suggesting stable binding of glucose at tryptophan-rich sites. Of the 18 tryptophan residues in Cwht1p, eight are exposed to the solvent, and four are found in

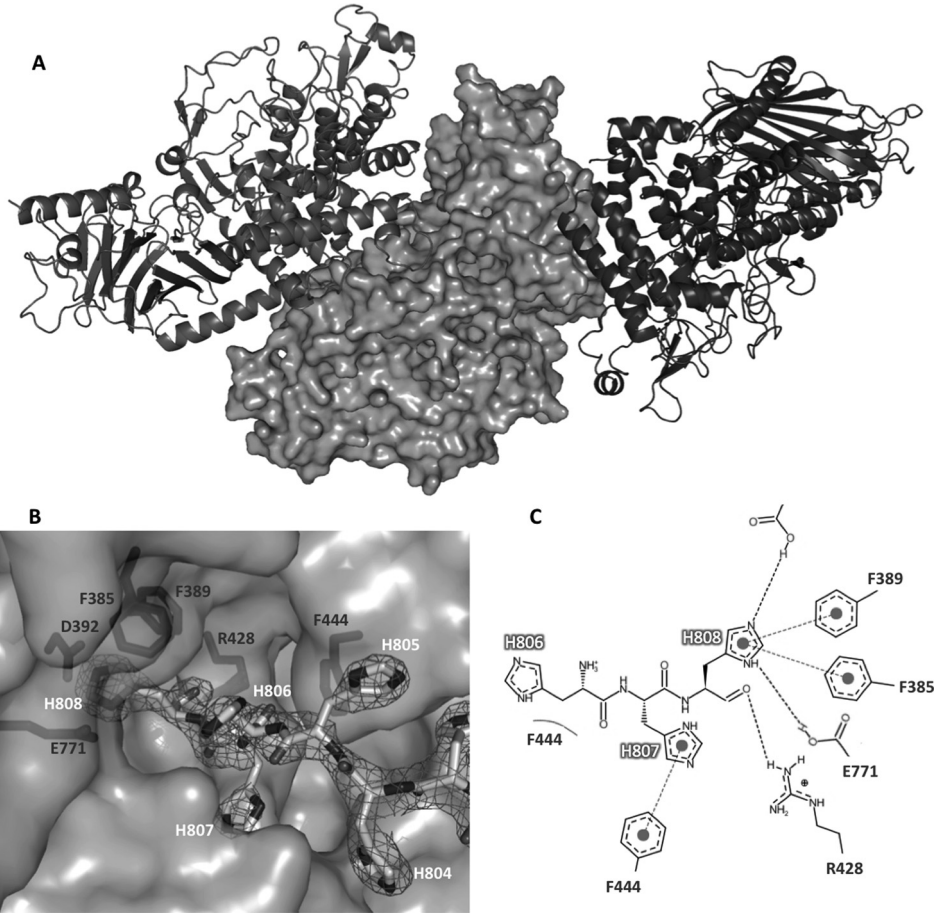

FIGURE 2. **Cwht1p packing in the crystal.** *A,* there are two major packing interfaces with the central *gray* monomer (surface representation), interacting with the lower ribbon structure (747 Å^2 buried) and upper ribbon structure (1456 Å^2 buried) of symmetry-related mates. *B,* the His_6 tag of the upper symmetry-related mate is found in the proposed active site cleft of the central monomer. The density around the histidine residues is shown in a $2F_o - F_c$ map contoured at 2.0 σ, and the residues are labeled in *white text*. Residues interacting with the His tag are shown in *sticks*, and labeled in *dark text*. *C,* two-dimensional representation of the interactions in the crystal between the terminal histidines of one monomer with its symmetry-related mate. Hydrogen bonds and stacking interactions shown in *dashed lines*.

TABLE 1

Structural alignment of $(\alpha/\alpha)_6$ domains of Cwht1p homologs

The $(\alpha/\alpha)_6$ domains of the top four characterized structural homologs from a DALI search were aligned with the C-terminal domain of Cwht1p using the DALI-Lite server.

Protein	PDB code	DALI score	R.m.s. deviations (Å) from Cwht1p C-domain	Sequence % ID with C-domain of Cwht1p	Catalytic residues	
E. coli GH63, YgjK	3D3I	30.7	2.5	20	Asp^{501a}	Glu^{727a}
Trehalase (Tre37A)	2JF4	30.1	2.6	15	Asp^{312}	Glu^{496}
Chitobiose phosphorylase (ChBP)	1V7W	25.3	3.2	11	Asp^{492}	Gln^{690}
Glucoamylase (tGA)	ILF6	20.5	3.4	12	Glu^{438}	Glu^{636}
Cwht1p					Asp^{568a}	Glu^{771a}

[a] Proposed residues.

the proposed active site pocket (Trp^{381}, Trp^{710}, Trp^{715}, and Trp^{789}). The presence of 18 tryptophans prevents conclusions about which particular residues may be interacting with the glucose molecules: multiple binding locations are possible, and glucose interaction could cause conformational changes potentially altering the local chemical environment around tryptophan residue(s). However, given the fact that half of the solvent-exposed tryptophans are found in the active site, and that this enzyme is active upon glucose-oligosaccharides giving glucose as a product, it is reasonable to expect that there is likely glucose binding to the active site, supported by the tryptophan fluorescence results. Thus, these data encourage further *in silico* docking to the Cwht1p active site for investigation of substrate/ligand binding modes.

Six known glycoside hydrolase inhibitors (supplemental Fig. S1) were screened for inhibition of Cwht1p activity, with results shown in Fig. 3*B*. Of the six, only miglitol was an effective inhibitor at 1 mM; a dose-response curve indicated that miglitol inhibits with an IC_{50} of 22 μM (Fig. 3*C*). Due to constraints on the tetrasaccharide substrate availability, a full K_i could not be determined; however, this IC_{50} is in the same range as the K_i value of the parent compound deoxynojirimycin (DNJM) of miglitol (50 μM). As a glucose analog, it is likely binding in the active site of the enzyme, as has been seen in other inhibitor-bound glycosidase structures (47–50).

In an attempt to displace the histidine tag from the active site, the structures of Cwht1p soaked with inhibitor (DNJM and

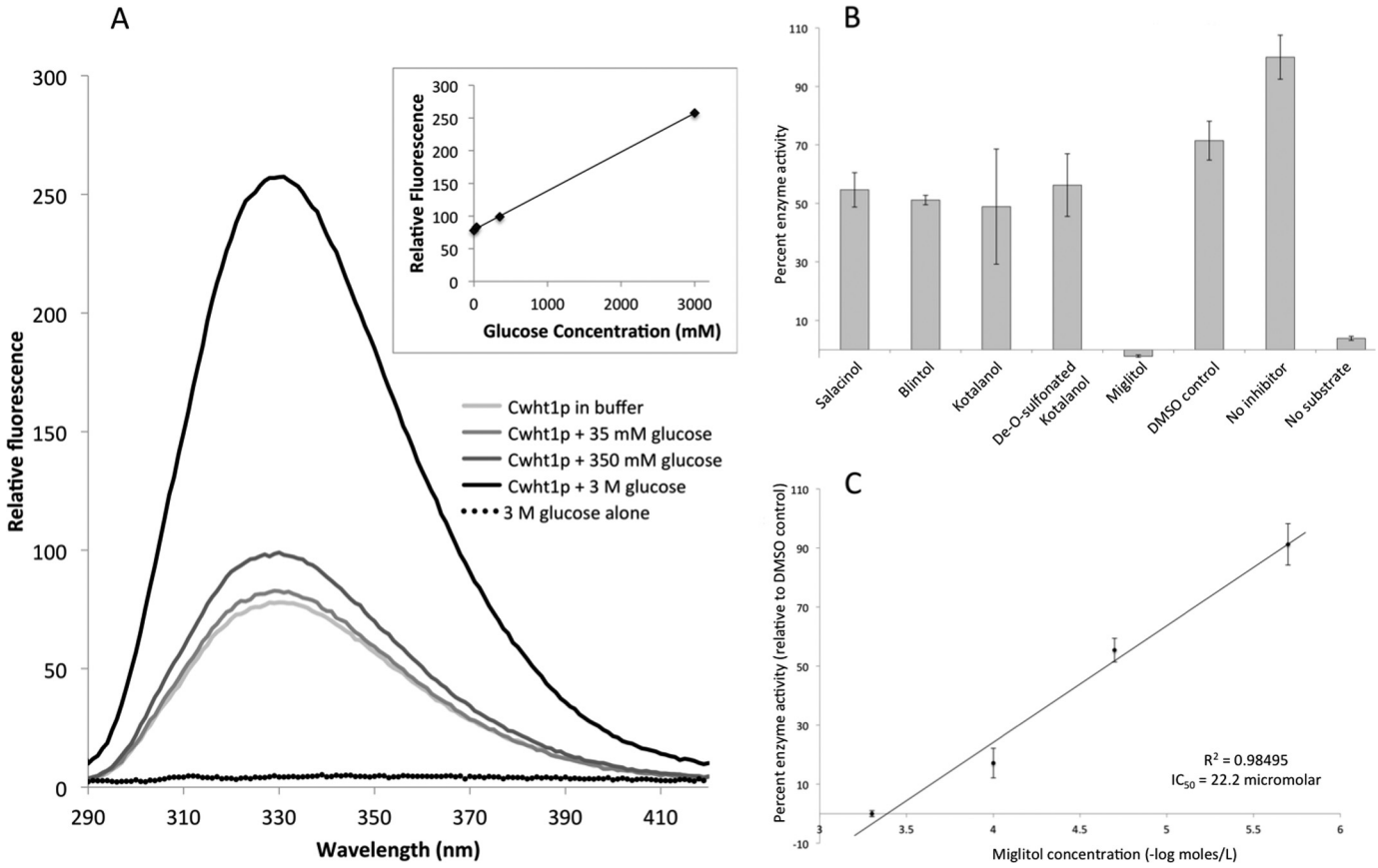

FIGURE 3. **Experimental support for glucose and inhibitors binding Cwht1p.** *A,* tryptophan fluorescence of Cwht1p in the presence and absence of glucose. Fluorescence increases upon the addition of glucose, but no peak shift is seen. *Inset*, relative fluorescence at the peak wavelength (330 nm). *B,* screening of compounds for inhibition of Cwht1p; all compounds are at a final 1 mM concentration. *C,* dose dependence of Cwht1p inhibition by miglitol.

miglitol) were solved to 2.1 Å (supplemental Table S3). Despite high concentrations of inhibitor, the histidine tag was not displaced from the active site of the symmetry-related mate (supplemental Fig. S6). Extensive efforts to crystallize the protein following enzymatic cleavage of the histidine tag were unsuccessful.

Glucose and Inhibitor Docking in Silico—The compounds α-D-glucose, miglitol, DNJM, and kojibiose (α-D-glucose-(1→2)-α-D-glucose) were independently docked to a single molecule of Cwht1p (that is, with no His tag in the active site) in a box around the active site, and the top binding results (poses) were modeled. The top poses for each are overlaid in the Cwht1p structure in Fig. 4*A*. The single-ring ligands all docked into two locations: the pocket containing the proposed catalytic residues ("site A"), and a second pocket roughly 12 Å away formed by residues 419–455 in the C′ region ("site B"). Site A is the proposed active site pocket, and the single-ring ligands docked here (including the top pose) make polar contacts with Cwht1p residues Trp^{391}, Asp^{392}, Arg^{428}, Gly^{566}, Asp^{568}, Trp^{710}, and Glu^{771}. The single-ring ligands binding to site B make polar contacts with Cwht1p residues Glu^{361}, Glu^{443}, Arg^{428}, Glu^{429}, Phe^{444}, Val^{446}, Gln^{447}, and Asn^{448}. No stacking interactions were seen between these ligands and Cwht1p aromatic residues. The calculated binding affinities for glucose, DNJM, and miglitol ranged from −5.2 to −6.0 kcal/mol, −4.7 to −5.2 kcal/mol, and −5.0 to −5.5 kcal/mol, respectively.

The top binding affinities for glucose, DNJM, and miglitol from the docking results were −5.9, −5.5, and −5.2 kcal/mol, which equate to calculated K_i values at 37 °C of 69, 132, and 215 μM, respectively. Glucose as an inhibitor has not been directly tested in the activity assay used here, as it is a substrate for the secondary/reporter glucose oxidase reaction. An indirect evaluation of glucose inhibition can come from its presence as a product of the Cwht1p catalytic reaction. In the activity assay, the reaction rate first decreases slightly at the longest time point tested (120 min), which contains 120 μM glucose product (38). This activity decrease is likely due to product inhibition of Cwht1p, and is seen to be in the same order of magnitude as the calculated glucose K_i value of 69 μM.

The disaccharide kojibiose contains the structure of the two terminal α(1→2)-linked glucoses in the natural substrate, $Glc_3Man_9GlcNAc_2$. As a larger molecule, kojibiose had a more complex set of results, with the top nine poses overlaid in the bottom panel of Fig. 4*A*. The top kojibiose poses ranged in calculated affinity from −6.5 to −7.1 kcal/mol. Four poses were found within site A, and overlaid well (maximum r.m.s. deviations 0.5 Å from top pose in this site). They are oriented with their non-reducing glucose in site A and their reducing end situated under the loop containing helix C′H1. Five poses were found in site B, four of which were situated with the non-reducing glucose of kojibiose in the site B pocket. These poses are loosely in the same position (maximum r.m.s. deviations 1.9 Å

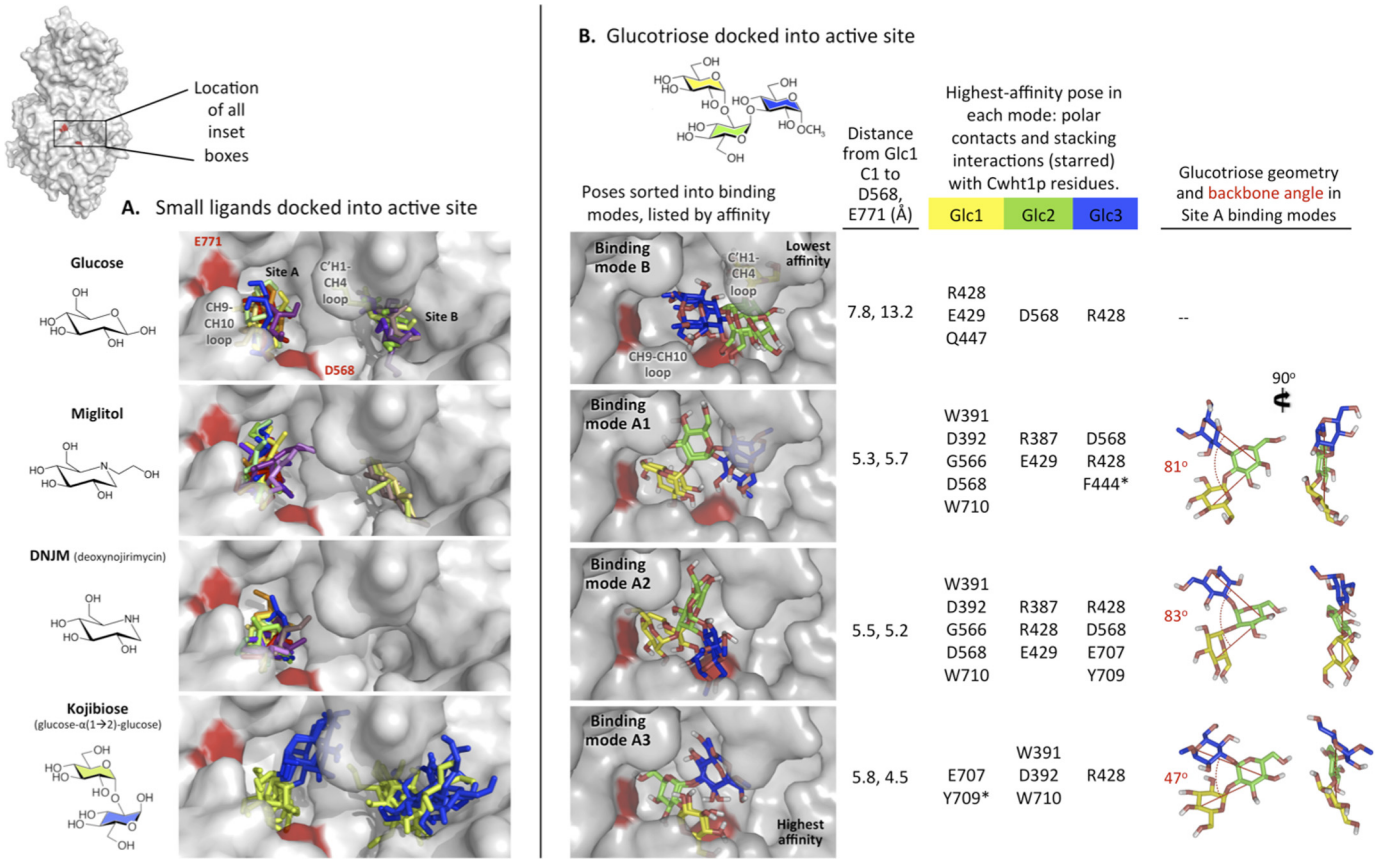

FIGURE 4. **Ligand and substrate docking to Cwht1p.** A surface representation of Cwht1p, in the same orientation as Fig. 1, with the active site residues indicated in *red*. Small ligands (glucose, DNJM, Miglitol, and kojibiose) and glucotriose were independently docked into a volume surrounding the active site, and all poses generated fell within the box drawn, into site A and/or site B. The *top* poses of each ligand are presented as *sticks*. *A,* small ligand binding. The calculated affinity of the poses is indicated in a rainbow: highest affinity in *red*, lowest in *violet*. The residues of each kojibiose molecule are shown in *green* (non-reducing terminus) and *yellow* (reducing terminus). *B*, glucotriose binding. The chemical structure of glucotriose is shown to indicate coloring. *Left*, surface of Cwht1p active site cleft with all top glucotriose poses shown, separated into four binding modes. To show the full sugar molecule in the active site, the model is rotated 45° to the right relative to *panel A*. The highest affinity binding mode is shown at *bottom*. *Center left*, distance from Glc1 Carbon 1 (anomeric carbon) to the proposed catalytic residues. *Center right*, polar and stacking contacts between the sugar residues of the top pose of each mode with Cwht1p amino acids. *Far right,* geometry of glucotriose poses in site A. The Glc1 (C4)-Glc2 (C4)-Glc3 (O5) angle is measured to compare the relative compactness of the trisaccharide geometries. The trisaccharide is shown with the angle in the plane of the page and normal to the page. The angle size is indicated in *red*.

between poses), all oriented with the non-reducing glucose in the pocket, and the reducing end pointing outward in a direction away from site A. The outlying pose found in site B has an opposite orientation: the reducing glucose is in the site B pocket, with the non-reducing glucose pointing outwards. No stacking interactions between kojibiose residues and Cwht1p aromatic residues were seen in any binding modes.

Glucotriose Docking in Silico—Glucotriose (α-D-glucose-(1→2)-α-D-glucose-α-D-(1→3)-[1-methyl-α-D-glucose]) is the minimum substrate cleaved by GluI. In this trisaccharide, glucoses 1, 2, and 3 are numbered from the non-reducing end. In docking results with respect to Cwht1p, the location of Glc1 is termed subsite −1, and Glc2 is found in subsite +1; Cwht1p cleaves the bond between these sugar residues. The electrophile of the glycoside hydrolysis reaction is carbon 1 of Glc1.

Glucotriose docking was performed to investigate potential catalytic binding sites of the minimum substrate of Cwht1p, and the subsites occupied by its sugar residues. The top nine docked poses were investigated, showing binding affinity ranging from −6.5 to −7.6 kcal/mol. These poses were manually sorted into sets of binding modes from the roughly common orientations between poses (Fig. 4*B*), giving four major binding modes: A1, A2, A3, and B, based upon the site (A or B) occupied by the trisaccharide. The Cwht1p contacts with each sugar residue are listed in Fig. 4*B*.

Of the four glucotriose binding modes, mode B is the most distinct. As seen in Fig. 4*B*, mode B contains Glc1 in site B, inaccessible to the proposed catalytic residues; there are no other suitable acid/base amino acids in this location to act as catalytic residues. As a result, this mode does not likely reflect the binding site for the glucotriose substrate, and it will be excluded from further analysis in proposing the catalytic site.

The remaining binding poses were docked into site A. In modes A1 and A2, Glc1 is found in the active site pocket. Conversely, in mode A3, the active site pocket contains Glc2. Comparing the A1 and A2 modes, the Glc2 and Glc3 residues are rotated 45° from the Glc1-Glc2 axis. In all three of these binding modes, carbon 1 of Glc1 is situated appropriately for glycoside hydrolysis by Asp^{568} and Glu^{771}, the proposed catalytic residues.

Determination of the Substrate-binding Model—In evaluation of the binding modes to determine the likeliest substrate-binding model, we investigated the glucotriose conformation. GluI enzymes from yeast and mammalian sources do not cleave

simple sugar substrates such as *p*-nitrophenyl-α-glucose; they require a minimum trisaccharide of three non-reducing terminal sugars from the 14-mer oligosaccharide substrate (22, 28, 38). This linkage is unique in biology; no other reports of such a glucotriose are found in the literature. As a result of Glc2 having two glycoside bonds at neighboring carbons (1 and 2), this glucotriose has a unique shape: it is non-linear, and bent back on itself, with an expected intra-chain interaction between Glc1 and Glc3. GluI is highly specific for this sugar, and does not cleave other linear glucose chains, indicating that this bent-back shape may be important for interaction with the active site, and may be a selectivity determinant for binding and/or catalysis.

The sugar conformation in binding modes A1 and A2 have similar shapes: they have an 80° angle along the chain, in comparison to the much more acute A3 mode (Fig. 4*C*). Furthermore, there is an intra-saccharide stacking interaction in A3, between Glc1 and Glc3, which is not seen in the other modes. This stacking interaction lines up well with Tyr^{709} in the active site. This evidence, with the highly specific and unique relationship between Cwht1p and its substrate, places binding mode A3 as the likeliest binding mode.

Residue Conservation Supports Mode A3 as the Substrate-binding Model—All eukaryotic GluI homologs share similar substrate specificity to the glucotriose discussed here. This substrate has not been tested with prokaryotic GluI homologs, which are active against other oligosaccharides; however, prokaryotic glycans have not been shown to contain the glucotriose oligosaccharide with the α(1→2) and α(1→3) linkages (51). Therefore, residue conservation in GH 63 enzymes can be considered in light of the binding model proposed here. Using the structure-based sequence alignment shown in supplemental Fig. S3, supplemental Table S3 presents the conservation of the key residues involved in mode A3, the likeliest binding model.

DISCUSSION

We have determined the first eukaryotic structure of GluI, a large glycosylated GH 63 enzyme responsible for the first step in the *N*-glycosylation trimming pathway. The catalytic residues are identified from the structure to be Asp^{568} and Glu^{771}, mutations of which abolish activity. Using docking methods with inhibitors and the substrate as ligands, we have mapped the active site cleft, and proposed a substrate binding model.

Cwht1p Structural Features and Key Residues—Cwht1p consists of two domains joined by a linker helix; the N-domain is a super-β-sandwich, and the C-domain is an $(\alpha/\alpha)_6$ toroid with additional structural units, termed the C′-region, on one face. The two closest structural neighbors (YgjK and TTHA00978) are GH Family 63 members with Cwh41p, validating the CAZy sequence-based classification for these proteins. Like Cwht1p, all of the characterized closest structural neighbors are inverting glycosidases (supplemental Table S2); the active sites of four of these nearest neighbors have been characterized. These enzymes vary in substrate specificity, and are similar to either both or one (the $(\alpha/\alpha)_6$ toroid) domain of Cwht1p. Catalytic activity is found in the center of the $(\alpha/\alpha)_6$ barrel. The substrates can access the active site on the same face as the C′ region; thus this region, which varies greatly between structural neighbors, likely provides substrate selectivity.

The role(s) of the super-β-sandwich domains in Cwht1p and its structural neighbors is unclear; this fold resembles a family of carbohydrate-binding molecules (52). It may be involved in protein-protein interaction with neighboring enzymes (oligosaccharyl transferase or α-glucosidase II) along the *N*-glycosylation trimming pathway, or in interaction with substrate *N*-glycoproteins. The glycans present on the *N*-domain of Cwht1p may or may not be found in the native enzyme; host cell and expression conditions often result in altered glycosylation patterns (53–57). The glycosylated status of Cwht1p does not affect catalytic activity *in vitro* (58). Thus, no specific conclusion can be made about the role or presence of these glycans in GluI across species.

Cwht1p is the soluble C-terminal construct of Cwh41p, a type II membrane protein. In the model of Cwht1p, the N terminus protrudes from the convex face of the protein. In Cwh41p, this would be the location of the 33-residue transmembrane pass across the endoplasmic reticulum membrane. There is no clear electrostatic or hydrophobic patch on this face of the protein to indicate interaction with the membrane. In previous overexpression studies of the full-length Cwh41p in *S. cerevisiae*, both the membrane-bound and a soluble truncated form were isolated, despite the presence of a range of protease inhibitors during purification (23). This evidence of proteolytic cleavage of the protein in its native host could indicate that a portion of Cwh41p is present and active in its soluble form in the endoplasmic reticulum, without being tethered to the membrane. Similar proteolytic release of transmembrane proteins has been seen for glycosyltransferases in the Golgi (59–60).

Previous studies have shown the *in vivo* and *in vitro* effects of mutations within α-glucosidase I. From the structure-based sequence alignment, residues associated with these mutations were aligned with the Cwht1p structure. The results are summarized in Table 2. In general, any mutations interfering with the active site have been shown to reduce activity, as expected. No *N*-domain mutations, nor any benign mutations, have been reported in these studies. Aside from Cwht1p, two smaller constructs of Cwh41p (Cwht2p and Cwht3p) were cloned and their expression attempted in *S. cerevisiae* by others (58) and in *P. pastoris* by ourselves (28). Analysis of the structure of Cwht1p sheds light on their lack of expression in the two different hosts (Table 2).

Catalytic Residues of GluI—Prior biochemical studies have proposed two pairs of possible catalytic residues in Cwh41p: Asp^{584} and Glu^{771} of Cwht1p (from primary sequence alignment with prokaryotic GluI, Ygjk (27)); and Asp^{584} and Glu^{580} (from primary sequence alignment with proposed mammalian GluI substrate binding motif (26)). Alanine mutations of Asp^{584} and Glu^{580} have shown a loss of Cwht1p activity (58). Within the structure solved here, these two residues are found at the N terminus of helix CH6. They face the interior of the structure, and make polar contacts with several residues in the loop following C′H2. Given the solvent-inaccessibility of these amino acids, they are not likely to be catalytic residues. However, mutation of these residues to alanine, a small nonpolar resi-

TABLE 2

Proposed effects of known α-glucosidase I mutations or truncations

Non-yeast mutations have been aligned to Cwht1p using the structure-based sequence alignment, shown in Figure S4. Structural terminology is consistent with Fig. 1 and supplemental Fig. S2.

Source	Mutation	Known effect	Aligned Cwht1p residue	Location on Cwht1p structure	Proposed explanation for mutation effect
Human, CDGiiB	R486T	No binding to Sepharose; no catalytic activity	Arg^{428}	Active site; in between sites A and B.	Improper binding pocket shape and hydrogen bonding atoms; reduction of substrate binding
Human, CDGiiB	F652L	No catalytic activity; reduced binding to sepharose.	Tyr^{635}	Loop between CH7 and C′S5	Restructuring of loop at active site; possible interference with substrate binding.
Yeast	G725R	Loss of activity	Gly^{692}	Loop between CH9 and CH10; facing protein interior	Large volume of arginine causes steric clashing and loop rearrangement.
Human, Der7–1 mutant	S440F	Loss of activity	Ser^{378}	On protein surface before CH2	Creation of a hydrophobic patch, prone to aggregation
Yeast, laboratory construct Cwht2p	C-terminal construct beginning at Arg^{287}	No expression		Construct begins at random coil linker region prior to LH1	Exposure of LH1 hydrophobic patch, conducive to misfolding and degradation or aggregation
Yeast, laboratory construct Cwht3p	C-terminal construct beginning at Met^{493}	No expression		Construct begins at long disordered loop between CH4 and CH5	Construct does not encode a full domain, resulting in incorrect folding

due, would disrupt the contacts with the C′-region, possibly leading to misfolding. This is consistent with the decreased expression levels and abrogated catalytic activity of the D584A and E580A constructs. No studies have been published documenting the mutation of Glu^{771} or its corresponding residue in other homologs. In the Cwht1p structure, Glu^{771} is solvent-accessible.

Alignment of the C-domain of Cwht1p with the top-ranked structural homologs reveals tight structural conservation with GH-G and GH-L fold clans, particularly of the $(\alpha/\alpha)_6$ barrel, despite the relatively low sequence identity (supplemental Table S2). The C′-domain is much more variable between structural homologs, consistent with their variable substrate specificities. Within the GH-G and GH-L clans, the active site is found at the center of the $(\alpha/\alpha)_6$ bundle. Hydrolysis proceeds via an acid-base mechanism, utilizing a pair of carboxylic acidic residues to catalyze the reaction. The catalytic residues (glutamate and aspartate) of the characterized structural neighbors align with Glu^{771} and Asp^{568} of Cwht1p, at the core of the bundle (Table 1 and supplemental Fig. S4*C*). Thus, these two amino acids are proposed to be the catalytic residues for glycoside hydrolysis of Cwh41p. Cwht1p single and double point mutants at Asp^{568} and Glu^{771} were expressed and purified (supplemental Fig. S5). They share similar expression and purification properties to the native Cwht1p, and circular dichroism data indicates that the mutations have not induced large structural deviations from the native state. However, they are unable to cleave the tetrasaccharide substrate. Thus we have obtained properly folded, non-catalytic mutants of Cwht1p for use in active site investigations.

Interestingly, in a structural overlay, the proposed catalytic residues Asp^{568} and Glu^{771} in Cwht1p align, respectively, with Asp^{501} and Glu^{727} of YgjK (Table 1 and supplemental Fig. S4*C*). This contrasts the primary sequence alignments (27), which supported Cwht1p Asp^{584} and Glu^{771} as catalytic residues. Thus, the initial primary sequence alignments were incorrectly aligned in the region of Cwht1p Asp^{568}. Three-dimensional information allows an improved structure-based sequence alignment of the yeast, bacterial, and mammalian GH 63 members (supplemental Fig. S3). Using this alignment, the previously proposed (26) mammalian (human) binding sequence ^{594}ERHLDLRCW^{602} aligns with Cwht1p ^{580}ELNVDALAW^{588}; this is found in the C′-region of the Cwht1p structure. This loop does not align well with the prokaryotic structures (YgjK and TTHA00978) and varies largely between similar structures. The mammalian binding sequence was proposed based upon chemical modification studies supporting the presence of Arg, Trp, Tyr, and Cys in the active site. However, there are several Arg, Trp, Tyr, and Cys residues in the three-dimensional Cwht1p structure that are conserved distant in primary sequence. In particular, Arg^{387}, Trp^{391}, Tyr^{709}, Trp^{710}, Arg^{711}, and Trp^{789} line the proposed binding pocket, and are conserved between the yeast and mammalian sequences in the structure-based alignment. Thus, the chemical modification experiments proposed the mammalian binding motif can be re-interpreted in light of the solved Cwht1p structure, to support the catalytic site surrounding Asp^{568} and Glu^{771}.

Crystal Packing and Utility of this Model for More Active Site Investigations—The crystal form used here (28) was the only reproducible form found from screening and optimizing sparse matrix screens. Packing analysis of this form shows that Cwht1p crystallized with two main interfaces and a total buried surface area of 2203 Å^2. Each of the interfaces is much larger and more thermodynamically favorable than average in crystallization (61, 62). The thermodynamic favorability of this crystal form could explain why it was seen in multiple distinct conditions, and why no other reproducible forms were seen with this construct across many conditions.

The large interface contains many interactions between the C-terminal His_6 tag of one monomer and the interior of the $(\alpha/\alpha)_6$ bundle of a crystallographic symmetry-related monomer (Fig. 2, *B* and *C*). These interactions contribute approximately one-fourth of the buried surface area and solvation free energy of that interface, a significant contribution from this non-native tag. Notably, His^{808} is hydrogen-bonded to Glu^{771}, a proposed catalytic residue. $His^{806-808}$ are also interacting with several highly conserved residues (Asp^{392}, Phe^{389}, Phe^{385}, and Phe^{444}) and one moderately conserved residue (Arg^{428}).

Following inhibitor (DNJM and miglitol) soaks at a high inhibitor concentration (10–100-fold K_i or IC_{50}), the histidine tag, rather than the inhibitor, was still clearly seen in the electron density of the active site (supplemental Fig. S6). Soaking

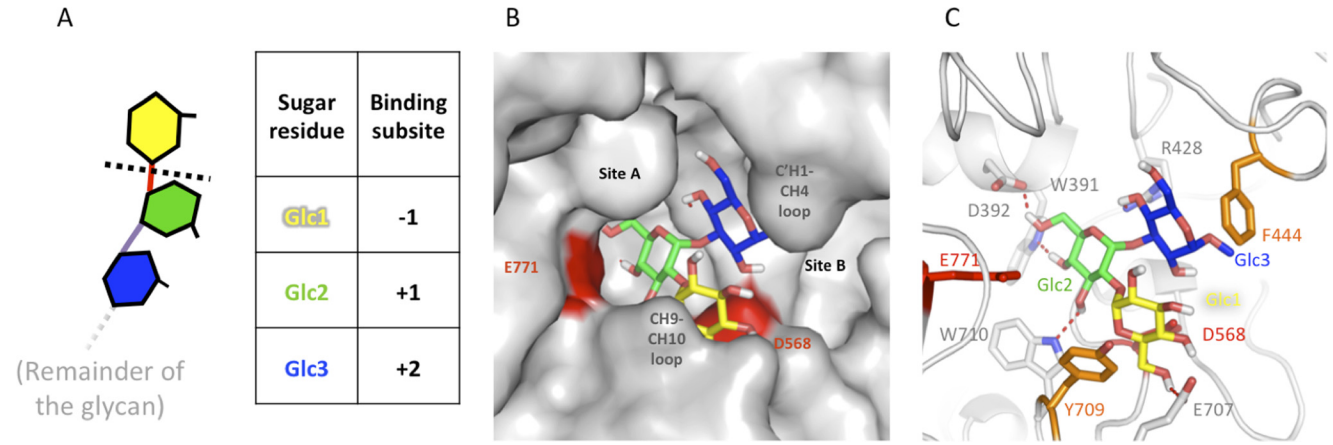

FIGURE 5. **Proposed model of substrate binding to the Cwht1p active site.** *A*, schematic of trisaccharide and subsite nomenclature. Colors are consistent with Fig. 4 glucotriose: the non-reducing terminal glucose is shown in *yellow*. Scissile bond is shown with a *dashed black line*. *B*, binding model shown on surface representation of Cwht1p; the molecule is oriented as described in the legend to Fig. 4*B*. Sites A and B are as determined from monosaccharide docking. Proposed sugar subsites −1, +1, and +2 are *circled in yellow*, *green*, and *blue*, respectively; catalytic residues are shown in *red*. In this model, the bond between Glc1 and Glc2 (the −1 and +1 subsites) is cleaved by Asp^{568} and Glu^{771}. *B*, polar contacts and stacking interactions between glucotriose and Cwht1p. Stacking interaction amino acids are shown in *orange*; proposed catalytic residues in *red*; polar contacts in *red dashed lines*. Glc3, Glc1, and Tyr^{709} stack together in the docking results. Phe^{444} is found on a flexible loop between helices C′H1 and CH4 that could move to stack with Glc3.

higher inhibitor concentrations resulted in crystal damage; it is likely that displacement of the histidine tag disrupts this major interface. As a result, this crystal form is not optimal for crystal soaks and active site mapping, and so we proceeded with an *in silico* approach to address these investigations.

Mapping Active Site with Inhibitors and Glucose—Intrinsic tryptophan fluorescence experiments here provide preliminary qualitative support for glucose binding (Fig. 3*A*). Additionally, three GluI inhibitors have been characterized to date: miglitol and DNJM, both single-ring glucose analogues, and the disaccharide kojibiose, α(1→2)-linked glucobiose. In this work, docking studies with these ligands were used to map the catalytic site.

All the single-ring ligands docked to two sites in the center of the Cwht1p catalytic domain: the proposed active site pocket (site A) and a nearby pocket, roughly 12 Å away (site B) (Fig. 4*A*). All top hits were found in site A. Within both sites, the ligands made polar contacts with several Cwht1p residues. Notably among the site A contacts, the ligands interact with two tryptophan residues (Trp^{710} and Trp^{391}), and are blocking solvent accessibility to two others (Trp^{715} and Trp^{789}); this result is consistent with our experimental tryptophan fluorescence data. The docked ligand molecules also interact with Asp^{568} and/or Glu^{771} (variable between poses), the proposed catalytic residues. Blocking substrate accessibility to these residues would certainly abrogate catalytic activity, and has been structurally seen in other glycoside hydrolases inhibited by monosaccharide analogs (47–50). Site B does not possess any carboxylic acid residues necessary for glycoside hydrolysis, and so is unlikely to be the active site pocket for cleavage of the terminal glucose from the 14-mer oligosaccharide substrate.

Interestingly, the top binding mode found kojibiose with the non-reducing glucose in site B, distal to the catalytic residues in the active site pocket. Kojibiose makes several polar contacts in site B; this pocket contains few hydrophobic residues, and so could be involved in binding the hydroxyl groups of non-terminal residues in the 14-mer biological substrate, $Glc_3Man_9GlcNAc_2$. Cwht1p mutagenesis experiments, kinetic evaluation, and co-crystallization are required to experimentally investigate the potential role of this pocket in inhibitor binding.

Glucotriose Structure in Determining the Substrate-binding Model—Based upon catalytic residue accessibility, GluI substrate selectivity, and glucotriose conformation, we evaluated the docked binding modes to propose a substrate binding model, as shown in Fig. 5. Subsite −1 is found under the loop between CH9 and CH10, with Glc1 stabilized by polar contact with Glu^{707} and a stacking interaction with Tyr^{709}. Subsite +1 is the "active site pocket" where Glc2 makes specific polar contacts with Trp^{391}, Asp^{392}, and Trp^{710} in this otherwise hydrophobic cleft. Subsite +2 contains Glc3, under the loop containing helix C′H1; Arg^{428} makes a polar contact with the sugar. Another sugar-protein contact could take place following a small conformational change from this pose: residue Phe^{444} is found in the C′H1 loop, and could interact with the Tyr^{709}-Glc1-Glc3 stacking. The anomeric carbon of this Glc3 points toward site B; the remainder of the 14-mer sugar could be found in this cleft or instead could protrude outwards from Cwht1p, making minimal protein contact. The latter is supported by the unchanged kinetic parameters between the trisaccharide and tetrasaccharide substrates (28, 45).

Cwht1p is expected to be a suitable model for other eukaryotic homologs, all of which share similar substrate specificity to the glucotriose oligosaccharide containing α(1→2) and α(1→3) linkages. Conserved residues involved in the proposed model are listed in supplemental Table S3. The catalytic residues Asp^{568} and Glu^{771} are conserved across all species investigated; similarly, all residues interacting with Glc2 are conserved. This is not surprising, as this is the binding site found in many $(\alpha/\alpha)_6$-barrel glycosidases with various oligosaccharide substrates, and the polar contacts here orient the sugar ring to place the anomeric carbon in place for glycoside bond cleavage by the catalytic residues (63–65). In prokaryotes, Tyr^{709} is highly conserved, but the residues interacting with Glc3 are not; in the prokaryotic structures, the area analogous to site B is occluded. Therefore, the unknown sugar substrate for the pro-

karyotic GH 63 members does not bind in a similar fashion as Glc3 in Cwht1p. However, eukaryotic conservation at both Glc1 and Glc3 sites supports the model proposed above in recognition of the glycan substrate by the yeast enzyme.

Validation of in Silico Work—Despite the use of non-flexible structure docking, the calculated K_i values for the single-residue ligands show reasonable agreement with the micromolar range of values determined experimentally, supporting the binding modes seen. Similarly, the kojibiose binding affinities (top mode 9.9 μM) were only roughly within the same order of magnitude as those determined experimentally (55 μM with membrane-bound Cwht1p (15)). However, in glucotriose docking, the calculated binding energy ranged from 52 to 71 μM affinity. This is 20-fold smaller than the experimental K_m value for the trisaccharide substrate with a hydrocarbon tail of 1.28 mM with Cwht1p (45). This deviation is not unexpected as the binding energies are more poorly predicted as the ligand size increases (66).

The precise structural location of inhibitors binding Cwht1p is not definitively known. However, the docked results here are in good agreement with single-ring ligands binding the active sites of other inverting glycosidases (47–50). In addition, inhibitor soak damage of this crystal form, heavily dependent on active site contacts, supports their binding at the active site, in accordance with the docking results. These observations regarding the accuracy of binding modes and affinities are in accordance with what is seen in the literature. The scored affinities are not highly accurate in general; the strength of docking methods lies largely in the accurate predictions of the pose orientations, and less so in energetic calculations (66, 67).

Conclusions—The structure presented here at 2-Å resolution, and its proposed substrate-binding model, establish the underlying basis for the high substrate specificity of eukaryotic GluI. Furthermore, these results demonstrate the use of *in silico* modeling as a method complementary to experimental work. The structure and model will inform further research into the relationship of GluI with its unique glucotriose substrate, and pave the way for investigation into structure-based drug design toward specific *N*-glycosylation inhibitors.

Acknowledgments—We acknowledge A. Chakrabartty and K. Popovic for assistance with circular dichroism and tryptophan fluorescence experiments; J. Holyoake and G. Li for helpful discussions regarding in silico experiments; and A. Beristain, for critical reading of the manuscript. C. Scaman provided the Cwh41p construct, and A. Fairbanks the tetrasaccharide substrate. Use of the Advanced Photon Source was supported by the United States Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. DE-AC02-06CH11357. Use of the BioCARS Sector 14 was also supported by grants from the National Center for Research Resources Grant 5P41RR007707 and the National Institutes of Health Grant 8P41GM103543 from the NIGMS. The Cornell High Energy Synchrotron Source (CHESS) is supported by the National Science Foundation and National Institutes of Health NIGMS via National Science Foundation Grant DMR-0936384; the MacCHESS resource is supported by National Institutes of Health Grant GM103485 from the NIGMS. Development of the HKL2000 software package is supported by National Institutes of Health Grant GM-53163 (to Z. Otwinowski and W. Minor).

REFERENCES

1. Apweiler, R., Hermjakob, H., and Sharon, N. (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta* **1473,** 4–8
2. Vigerust, D. J., and Shepherd, V. L. (2007) Virus glycosylation. Role in virulence and immune interactions. *Trends Microbiol.* **15,** 211–218
3. Roth, J., Zuber, C., Park, S., Jang, I., Lee, Y., Kysela, K. G., Le Fourn, V., Santimaria, R., Guhl, B., and Cho, J. W. (2010) Protein *N*-glycosylation, protein folding, and protein quality control. *Mol. Cells* **30,** 497–506
4. Helenius, A., and Aebi, M. (2004) Roles of *N*-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **73,** 1019–1049
5. Burda, P., and Aebi, M. (1998) The ALG10 locus of *Saccharomyces cerevisiae* encodes the α-1,2 glucosyltransferase of the endoplasmic reticulum. The terminal glucose of the lipid-linked oligosaccharide is required for efficient *N*-linked glycosylation. *Glycobiology* **8,** 455–462
6. Spiro, R. G. (2000) Glucose residues as key determinants in the biosynthesis and quality control of glycoproteins with *N*-linked oligosaccharides. *J. Biol. Chem.* **275,** 35657–35660
7. Spiro, M. J., and Spiro, R. G. (1991) Potential regulation of *N*-glycosylation precursor through oligosaccharide-lipid hydrolase action and glucosyltransferase-glucosidase shuttle. *J. Biol. Chem.* **266,** 5311–5317
8. Durrant, C., and Moore, S. E. (2002) Perturbation of free oligosaccharide trafficking in endoplasmic reticulum glucosidase I-deficient and castanospermine-treated cells. *Biochem. J.* **365,** 239–247
9. De Praeter, C. M., Gerwig, G. J., Bause, E., Nuytinck, L. K., Vliegenthart, J. F., Breuer, W., Kamerling, J. P., Espeel, M. F., Martin, J. J., De Paepe, A. M., Chan, N. W., Dacremont, G. A., and Van Coster, R. N. (2000) A novel disorder caused by defective biosynthesis of *N*-linked oligosaccharides due to glucosidase I deficiency. *Am. J. Hum. Genet.* **66,** 1744–1756
10. Mehta, A., Zitzmann, N., Rudd, P. M., Block, T. M., and Dwek, R. A. (1998) α-Glucosidase inhibitors as potential broad based antiviral agents. *FEBS Lett.* **430,** 17–22
11. Durantel, D., and Alotte, C. (2007) Glucosidase inhibitors as antiviral agents for hepatitis B and C. *Curr. Opin. Investig. Drugs* **8,** 125–129
12. Chapel, C., Garcia, C., Bartosch, B., Roingeard, P., Zitzmann, N., Cosset, F. L., Dubuisson, J., Dwek, R. A., Trépo, C., Zoulim, F., and Durantel, D. (2007) Reduction of the infectivity of hepatitis C virus pseudoparticles by incorporation of misfolded glycoproteins induced by glucosidase inhibitors. *J. Gen. Virol.* **88,** 1133–1143
13. Asano, N., Oseki, K., Kizu, H., and Matsui, K. (1994) Nitrogen-in-the-ring pyranoses and furanoses. Structural basis of inhibition of mammalian glycosidases. *J. Med. Chem.* **37,** 3701–3706
14. Papandréou, M. J. (2002) The α-glucosidase inhibitor 1-deoxynojirimycin blocks human immunodeficiency virus envelope glycoprotein-mediated membrane fusion at the CXCR4 binding step. *Mol. Pharmacol.* **61,** 186–193
15. Bause, E., Schweden, J., Gross, A., and Orthen, B. (1989) Purification and characterization of trimming glucosidase I from pig liver. *Eur. J. Biochem.* **183,** 661–669
16. Shailubhai, K., Pukazhenthi, B. S., Saxena, E. S., Varma, G. M., and Vijay, I. K. (1991) Glucosidase I, a transmembrane endoplasmic reticular glycoprotein with a luminal catalytic domain. *J. Biol. Chem.* **266,** 16587–16593
17. Faridmoayer, A. (2004) An improved purification procedure for soluble processing α-glucosidase I from *Saccharomyces cerevisiae* overexpressing CWH41. *Protein Expr. Purif.* **33,** 11–18
18. Romero, P. A., Dijkgraaf, G. J., Shahinian, S., Herscovics, A., and Bussey, H. (1997) The yeast *CWH41* gene encodes glucosidase I. *Glycobiology* **7,** 997–1004
19. Bause, E., Erkens, R., Schweden, J., and Jaenicke, L. (1986) Purification and characterization of trimming glucosidase I from *Saccharomyces cerevisiae*. *FEBS Lett.* **206,** 208–212
20. Hettkamp, H., Legler, G., and Bause, E. (1984) Purification by affinity chromatography of glucosidase I, an endoplasmic reticulum hydrolase involved in the processing of asparagine-linked oligosaccharides. *Eur. J. Biochem.* **142,** 85–90
21. Shailubhai, K., Pratta, M. A., and Vijay, I. K. (1987) Purification and characterization of glucosidase I involved in *N*-linked glycoprotein processing in bovine mammary gland. *Biochem. J.* **247,** 555–562

22. Zeng, Y. C., and Elbein, A. D. (1998) Purification to homogeneity and properties of plant glucosidase I. *Arch. Biochem. Biophys.* **355,** 26–34
23. Dhanawansa, R., Faridmoayer, A., van der Merwe, G., Li, Y. X., and Scaman, C. H. (2002) Overexpression, purification, and partial characterization of *Saccharomyces cerevisiae* processing α glucosidase I. *Glycobiology* **12,** 229–234
24. Henrissat, B., and Bairoch, A. (1996) Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.* **316,** 695–696
25. Palcic, M. M., Scaman, C. H., Otter, A., Szpacenko, A., Romaniouk, A., Li, Y. X., and Vijay, I. K. (1999) Processing α-glucosidase I is an inverting glycosidase. *Glycoconj. J.* **16,** 351–355
26. Romaniouk, A., and Vijay, I. K. (1997) Structure-function relationships in glucosidase I. Amino acids involved in binding the substrate to the enzyme. *Glycobiology* **7,** 399–404
27. Kurakata, Y., Uechi, A., Yoshida, H., Kamitori, S., Sakano, Y., Nishikawa, A., and Tonozuka, T. (2008) Structural insights into the substrate specificity and function of *Escherichia coli* K12 YgjK, a glucosidase belonging to the glycoside hydrolase family 63. *J. Mol. Biol.* **381,** 116–128
28. Barker, M. K., Wilkinson, B. L., Faridmoayer, A., Scaman, C. H., Fairbanks, A. J., and Rose, D. R. (2011) Production and crystallization of processing α-glucosidase I. *Pichia pastoris* expression and a two-step purification toward structural determination. *Protein Expr. Purif.* **79,** 96–101
29. Schweden, J., Borgmann, C., Legler, G., and Bause, E. (1986) Characterization of calf liver glucosidase I and its inhibition by basic sugar analogs. *Arch. Biochem. Biophys.* **248,** 335–340
30. Pukazhenthi, B. S., Varma, G. M., and Vijay, I. K. (1993) Conserved structural features in glycoprotein processing glucosidase I from several tissues and species. *Indian J. Biochem. Biophys.* **30,** 333–340
31. Faridmoayer, A., and Scaman, C. H. (2005) Binding residues and catalytic domain of soluble *Saccharomyces cerevisiae* processing α-glucosidase I. *Glycobiology* **15,** 1341–1348
32. Boggon, T. J., and Shapiro, L. (2000) Screening for phasing atoms in protein crystallography. *Structure* **8,** R143-R149
33. Otwinowski, Z., and Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326
34. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H.(2010) PHENIX. A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221
35. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501
36. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53,** 240–255
37. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67,** 235–242
38. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity. All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 12–21
39. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res.* **28,** 235–242
40. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372,** 774–797
41. Holm, L., and Rosenström, P. (2010) Dali server. Conservation mapping in 3D. *Nucleic Acids Res.* **38,** W545–9
42. Pei, J., Kim, B. H., and Grishin, N. V. (2008) PROMALS3D. A tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36,** 2295–2300
43. Jo, S., Vargyas, M., Vasko-Szedlar, J., Roux, B., and Im, W. (2008) PBEQ-solver for online visualization of electrostatic potential of biomolecules. *Nucleic Acids Res.* **36,** W270–275
44. Stierand, K., Maass, P. C., and Rarey, M. (2006) Molecular complexes at a glance. Automated generation of two-dimensional complex diagrams. *Bioinformatics* **22,** 1710–1716
45. Neverova, I., Scaman, C. H., Srivastava, O. P., Szweda, R., Vijay, I. K., and Palcic, M. M. (1994) A spectrophotometric assay for glucosidase I. *Anal. Biochem.* **222,** 190–195
46. Trott, O., and Olson, A. J. (2010) AutoDock Vina. Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31,** 455–461
47. Lieberman, R. L., D'aquino, J. A., Ringe, D., and Petsko, G. A. (2009) Effects of pH and iminosugar pharmacological chaperones on lysosomal glycosidase structure and stability. *Biochemistry* **48,** 4816–4827
48. Gloster, T. M., Turkenburg, J. P., Potts, J. R., Henrissat, B., and Davies, G. J. (2008) Divergence of catalytic mechanism within a glycosidase family provides insight into evolution of carbohydrate metabolism by human gut flora. *Chem. Biol.* **15,** 1058–1067
49. Harris, E. M., Aleshin, A. E., Firsov, L. M., and Honzatko, R. B. (1993) Refined structure for the complex of 1-deoxynojirimycin with glucoamylase from *Aspergillus awamori* var. X100 to 2.4-Å resolution. *Biochemistry* **32,** 1618–1626
50. Aleshin, A. E., Feng, P. H., Honzatko, R. B., and Reilly, P. J. (2003) Crystal structure and evolution of a prokaryotic glucoamylase. *J. Mol. Biol.* **327,** 61–73
51. Dell, A., Galadari, A., Sastre, F., Hitchen, P. (2010) Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes. *Int. J. Microbiol.* **2010,** 148178
52. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Carbohydrate-binding modules. Fine-tuning polysaccharide recognition. *Biochem. J.* **382,** 769–781
53. Knezević, A., Gornik, O., Polasek, O., Pucic, M., Redzic, I., Novokmet, M., Rudd, P. M., Wright, A. F., Campbell, H., Rudan, I., and Lauc, G. (2010) Effects of aging, body mass index, plasma lipid profiles, and smoking on human plasma *N*-glycans. *Glycobiology* **20,** 959–969
54. Varki, A., Freeze, H., and Gagneux, P. (2009) in *Evolution of Glycan Diversity, Essentials of Glycobiology* (Varki, A., Cummings, R. D., and Esko, J. D., eds) pp. 281–292, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
55. Ohtsubo, K., and Marth, J. D. (2006) Glycosylation in cellular mechanisms of health and disease. *Cell* **126,** 855–867
56. Dennis, J. W., Granovsky, M., and Warren, C. E. (1999) Protein glycosylation in development and disease. *Bioessays* **21,** 412–421
57. Gagneux, P., and Varki, A. (1999) Evolutionary considerations in relating oligosaccharide diversity to biological function. *Glycobiology* **9,** 747–755
58. Faridmoayer, A., and Scaman, C. H. (2007) Truncations and functional carboxylic acid residues of yeast processing α-glucosidase I. *Glycoconj. J.* **24,** 429–437
59. Lammers, G., and Jamieson, J. C. (1989) Studies on the effect of lysosomotropic agents on the release of Galβ1–4GlcNAc α-2,6-sialytransferase from rat liver slices during the acute-phase response. *Biochem. J.* **261,** 389–393
60. Jaskiewicz, E., Zhu, G., Bassi, R., Darling, D. S., and Young, W. W. (1996) β1,4-*N*-Acetylgalactosaminyltransferase (GM2 synthase) is released from Golgi membranes as a neuraminidase-sensitive, disulfide-bonded dimer by a cathepsin D-like protease. *J. Biol. Chem.* **271,** 26395–26403
61. Carugo, O., and Argos, P. (1997) Protein-protein crystal-packing contacts. *Protein Sci.* **6,** 2261–2263
62. Vekilov, P. G. (2003) Solvent entropy effects in the formation of protein solid phases. *Methods Enzymol.* **368,** 84–105
63. Okuyama, M. (2011) Function and structure studies of GH family 31 and 97 α-glycosidases. *Biosci. Biotechnol. Biochem.* **75,** 2269–2277
64. Gibson, R. P., Gloster, T. M., Roberts, S., Warren, R. A., Storch de Gracia, I., García, A., Chiara, J. L., and Davies, G. J. (2007) Molecular basis for trehalase inhibition revealed by the structure of trehalase in complex with potent inhibitors. *Angew. Chem. Int. Ed. Engl.* **46,** 4115–4119
65. Quiocho, F. (1989) Protein-carbohydrate interactions. Basic molecular features. *Pure Appl. Chem.* **61,** 1293–1306
66. Plewczynski, D., Łaźniewski, M., Augustyniak, R., and Ginalski, K. (2011) Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. *J. Comput. Chem.* **32,** 742–755
67. Gohlke, H., and Klebe, G. (2002) Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem. Int. Ed. Engl.* **41,** 2644–2676